D2.

(19) Eur päisches Patentamt
Europ an Patent Office
Offic urop´ n des br vets

(11) EP 1 047 077 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
25.10.2000 Bulletin 2000/43

(51) Int Cl.7: **G11C 8/00**, G11C 16/06, G11C 7/18

(21) Application number: **99830236.8**

(22) Date of filing: **21.04.1999**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant: **STMicroelectronics S.r.l.**
20041 Agrate Brianza (Milano) (IT)

(72) Inventors:
- **Micheloni, Rino**
  22078 Turate (IT)
- **Campardo, Giovanni**
  24128 Bergamo (IT)
- **Commodaro, Stefano**
  20060 Cassina de' Pecchi (IT)
- **Farina, Francesco**
  63023 Fermo (IT)

(74) Representative: **Cerbaro, Elena, Dr. et al**
STUDIO TORTA S.r.l.,
Via Viotti, 9
10121 Torino (IT)

(54) Nonvolatile memory device with double hierarchical decoding

(57) The memory array (30) comprises a plurality of cells (50), grouped together in sectors (31) and arranged in sector rows and columns, and has both hierarchical row decoding and hierarchical column decoding. Global word lines (35) are connected to at least two word lines (36) in each sector (31), through local row decoders (33); global bit lines (42) are connected to at least two local bit lines (43) in each sector (31), through local column decoders (40). The global column decoder (41) is arranged in the centre of the memory array (30), and separates from each other an upper half (30a) and a lower half (30b) of the memory array (30). Sense amplifiers (47) are also arranged in the middle of the array, thus saving space. This architecture also provides lesser stress of the cells, better reliability, and better production performance. In addition, each sector (31) is completely disconnected from the remaining sectors, and only the faulty row or column of a single sector should be doubled.

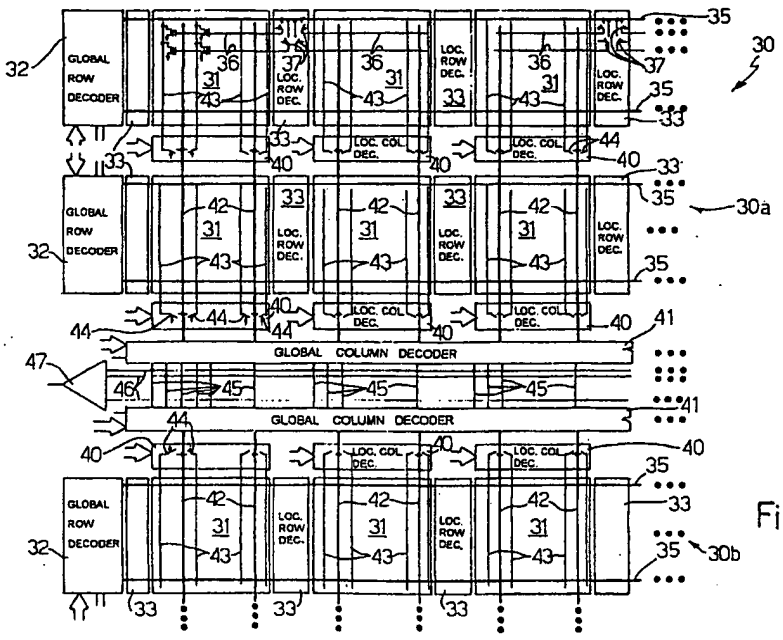

Fig.4

EP 1 047 077 A1

D scription

[0001] The present invention relates to a nonvolatile memory device, in particular a flash-EEPROM.

[0002] As known, memory arrays comprise a plurality of cells arranged on rows and columns, and each memory cell is identified by an own address, which, decoded, unambiguously addresses and biases the row and the column of the cell, so that the cell can be read and written. In general, the rows are known as word lines, and the columns are known as bit lines. In some cases a further subdivision in global lines and local lines is necessary; therefore local lines or columns connected directly to the cell to be addressed (to be read or written), are grouped in packets formed by a number of lines, usually a power of two, and are connected by respective selection circuits to a global line, addressed by a global decoder.

[0003] In general, standard decoding is an addressing system wherein the word or bit lines are of the same level to each other and are not organised hierarchically, and hierarchical decoding is an addressing system comprising global lines and local lines.

[0004] In flash memories, a further division into sectors is necessary (each sector being a block of cells having a common terminal and comprising an equal number of rows and columns), for which erasing is carried out. To avoid stress phenomena, detrimental to the cells, it is necessary to electrically isolate the various sectors to apply the high voltages (for example necessary for programming and erasing, and, for multi-level memories, also for reading), only to the selected sector. This division is dependent on the adopted erasing method, and two cases in particular can be distinguished:

- erasing by the source terminal, with the gate terminal grounded: in this case the sectors are organised by columns, and the row is shared. This solution has the disadvantage that the cells not to be programmed, but connected to the row of the cell to be written, are subjected to gate stress during the programming;
- erasing by applying a negative voltage to the gate region; in this case, two different decoding types are possible, i.e. hierarchical column decoding, and hierarchical row decoding.

[0005] Two examples of hierarchical (vertical) column decoding are shown in figures 1 and 2, wherein row decoding is standard and the column decoding is hierarchical.

[0006] In detail, figure 1 shows a memory array 1, comprising a plurality of sectors 2, arranged on two columns separated from each other by two pairs of row decod rs 3, for the right- and left-hand sectors respectively. A respective column decoder 4 is arranged below each sector 2. Each sector 2 comprises a plurality of memory cells 5, shown partially only for one s ctor 2.

[0007] In turn, the memory cells 5 are arranged on rows and columns; in particular the memory cells 5 arranged on a same row and belonging to a same sector 2 are connected to a same word line 6, which extends along the entire width of each sector, from the resp ctive row decoder 3. In addition, the memory cells 5 arranged on a same column and belonging to a same sector 2 are connected to a same local bit line 8, which extends along the entire height of each sector 2, from the respective local column decoder 4. Adjacent pairs of local bit lines 8 belonging to a same sector 2 are connected, through respective switches 10, for example formed by pass transistors, to a same global bit line 11; each global bit line extends along the entire respective sector column and is connected to pairs of local bit lines 8 belonging to vertically aligned sectors 2. The global bit lines 11 are connected to a global column decoding circuit 12, in turn connected to a reading stage 13 (comprising a plurality of sense amplifiers, not shown). The row decoders 3, the local column decoders 4, and the global column decoder 12, receive respective address and control signals, and the bias voltages required at respective inputs, as shown in figure 1.

[0008] The solution in figure 1 can be used if the number of sectors is not too high, and does not alter excessively the shape of the array, here an elongate rectangle. In case of a large number of sectors, it is possible to use the solution of figure 2.

[0009] Figure 2 shows a memory array 1', comprising a plurality of sectors 2, arranged on four columns and a plurality of rows. In particular, two row decoders 3 are arranged between each pair of sectors 2. In memory array 1', it is necessary to double the number of reading circuits, as shown in the figure by two reading stages 13. As an alternative, it is possible to introduce an additional multiplexing level. Otherwise, the memory array 1' in figure 2 is similar to those in figure 1.

[0010] In the hierarchical column decoding just described, the length of the rows and columns cannot be long, thus limiting the number and dimensions of the sectors to be provided. In fact, a long column can cause leakage, whereas the length of the row affects the access time, due to the equivalent time constant RC of the word lines. However, the solutions designed to solve this problem are affected in turn by further problems. For example, to reduce the problem of the biasing delay in case of long word lines, it has been proposed to arrange, in parallel with each polysilicon word line, a metal line, connected at various points to the actual word line (metal strap technique). This solution makes it possible to reduce the resistance of the word lines, but reduces the reliability of the memory, since it increases the risk of short-circuiting between the metal lines, which must be arranged along the path of the row, i.e. along its width.

[0011] In addition, in case of hierarchical column decoding, the problem exists that the row decoders occupy a large area, and the memory array as a whole becomes voluminous.

[0012] In case of hierarchical row decoding, the situation is inverse, i.e. the column decoding is standard, and row decoding is hierarchical. An embodiment of a memory array with hierarchical row decoding is shown in figure 3.

[0013] In detail, figure 3 shows a memory array 1" comprising a plurality of sectors 2, arranged on a plurality of rows and a plurality of columns, for example on four rows and eight columns, only some whereof are shown in figure 3. Each sector row is associated to a global row decoder 15, and each sector 2 is associated to a local row decoder 16, which is divided into two parts, arranged respectively to the left and right of each sector 2; in figure 3, the local row half-decoder 16, arranged to the right of a sector 2, forms a unit with the local row half-decoder 16 arranged to the left of the sector 2 adjacent on the right.

[0014] A plurality of global word lines 20 extends from each global row decoder 15 along the entire width of the memory array 1", i.e. along the entire row; each global word line 20, formed by a metal line, is thus connected to a plurality of polysilicon local word lines 21; in the example illustrated, for each sector 2, two local word lines 21 are connected to a same global word line 20, through a respective switch 22, for example comprising a pass transistor.

[0015] Each row of sectors 2 is associated to a respective column decoder 23, arranged alternately below and above the sector row; metal bit lines 24 extend from the column decoders 23 and are connected to the drain terminals of the memory cells 5. Adjacent column decoders 23 have first common output lines 25, connected to second output lines 26 and leading to a reading stage 27.

[0016] Similarly to the above, the global row decoders 15, the local row decoders 16 and the column decoders 23 receive respective address and control signals and the required bias voltages at the respective inputs, as shown in figure 3.

[0017] The hierarchical row decoding shown in figure 3 has disadvantages similar to those previously described for the hierarchical column decoding, i.e. the bit lines 24, the length of which is determined by the height (number of rows) of the sector, and the local word lines 21, the length of which is determined by the width (number of columns) of the sector, can be very long, and can cause leakages and long access times. Further disadvantages of this solution are because two channels (group of output lines 26) are present and convey the output signals to the reading stages 27, and four column decoders 23 are necessary. Consequently, this solution requires a large area. To eliminate this problem, it is possible to use a single central channel; however, here, the load of the bit lines 24 is doubled, which, in very large memories, can produce an unacceptable delay in loading and discharging the bit lines 24; in addition, leakages associated with the bit lines 24 increase, and the drain stress is greater. It is possible to reduce loading only suing a further column decoding level; however, this further level is critical, since it causes a greater voltage drop at the drain terminal of the cells 5. Another problem with hierarchical row decoding consists in the fact that the reading circuits 27 must be doubled, or require multiplexing.

[0018] The aim of the invention is to provide a different architecture for a nonvolatile memory, which eliminates the above described problems of the known solutions.

[0019] According to the present invention, a nonvolatile memory device is provided, as defined in claim 1.

[0020] To help understanding of the present invention, a preferred embodiment is now described, purely by way of non-limiting example, with reference to the attached drawings, wherein:

- figure 1 shows a first, known architecture of a memory array;
- figure 2 shows a second, known architecture of a memory array;
- figure 3 shows a third, known architecture of a memory array;
- figure 4 shows the architecture of a memory array according to the present invention; and
- figure 5 shows a cross-section through a memory cell belonging to the array of figure 4.

[0021] Figure 4 shows a memory array 30 according to the invention. The memory array 30 comprises a plurality of sectors 31 arranged on a plurality of rows and on a plurality of columns, for example on four rows and eight columns (only some of which are shown in figure 4). In turn, the sectors 31 comprise groups of memory cells 50 of flash-EEPROM type, having a common source terminal (here shown connected to ground). Each sector row is associated to a global row decoder 32 and each sector 31 is associated to a local row decoder 33; the row decoder 33 is divided into two parts, arranged respectively to the left and to the right of each sector 2; in figure 4, the local row half-decoder 33, arranged to the right of a sector 31, forms a unit with the local row half-decoder 33, arranged to the left of the sector 31 adjacent on the right.

[0022] A plurality of global word lines 35 extends from each global row decoder 32 along the entire width of the memory array 30, i.e. along the entire row; each global word line 35 is thus connected to a plurality of local word lines 36, two in the example illustrated, but typically four local word lines 36 for each sector 31, through respective switches 37, for example formed by pass transistors. As visible, the local word lines 36 connected to a same global word line 35 are aligned to each other on two rows.

[0023] Each sector 31 is associated to a respective local column decoder 40, in turn connected to a global column decoder 41, arranged horizontally in the center of the memory array 30. In detail, two pluralities of global bit lines 42 extend from the global column decoder 41,

shown divided into two parts for illustration, i.e. a first plurality extends from the global column decoder 41 towards the upper half of the memory array 30 (indicated at 30a), and a second plurality extends from the global column decoder 41 towards the lower half of the memory array 30 (indicated at 30b). Each global bit line 42 is connected to two local bit lines 43 for each sector 31, through respective switches 44, for example comprising pass transistors. As visible, the local bit lines 43 connected to a same global bit line 42 are aligned to each other on two columns. Similarly, the global bit lines 42 extending towards the upper half 30a and towards the lower half 30b of the memory array 30 are aligned to each other, and the respective local bit lines 43 are aligned along the columns of the memory array.

[0024] The global column decoder 41 has first output lines 45 connected to second output lines 46, leading to a reading stage 47, comprising a plurality of reading circuits (sense amplifiers), not shown, in a known manner.

[0025] As shown schematically in figure 4, the global row decoders 32, the local row decoders 33, the global column decoder 41 and the local column decoders 40 receive respective address and control signals, and the required bias voltage at the respective inputs.

[0026] For manufacturing the memory 30, a process with three metal layers is used, as shown schematically in figure 5, showing a cross-section of a memory cell 50. In detail, the cell 50 comprises a drain region 51 and a source region 52, housed in a substrate 53. A floating gate region 54 is formed above the substrate 53 and is insulated from the substrate 53 by a gate oxide layer 55; a polysilicon layer 57 extends above the floating gate region 54 and isolated therefrom by a dielectric layer 56, the polysilicon layer 57 forms a control gate region for the cell 50 and a local word line 36 (figure 4). A first dielectric layer 60 covers the substrate 53 and the polysilicon line 57. A first metal line 61, forming a local bit line 43 (figure 4), extends above the first dielectric layer 60, and electrically contacts the drain region 51. A second dielectric layer 62 extends above the first dielectric layer 60 and the first metal line 61. A second metal line 64, visible only in cross-section in figure 5 and forming a global word line 35 (figure 4), extends above the second dielectric layer 62, and is electrically connected to the local word line 36 through a switch 37 (not shown in figure 5, as schematically illustrated in figure 4). A third dielectric layer 65 extends above the second dielectric layer 62 and the second metal line 64. A third metal line 67, forming a global bit line 42, extends above the third dielectric layer 65 and is electrically connected to the first metal line 61 through a switch 44 (not shown in figure 5, as schematically illustrated in figure 4). A passivation layer 70 covers the entire device.

[0027] The architecture in figure 4 thus uses double hierarchical decoding, both of rows and columns. This solution has many advantages. First, th architecture is more compact, sinc cell reading requires only a single, central channel; this arrangement reduces th bulk and this reduction is only partially balanced by the need to provide local column decoding circuits.

[0028] The reading stages 47 can be arranged in the central area of the memory array 30, instead of to the right or to the left of the latter, thus contributing to reduction of area.

[0029] The memory array 30 is subjected to less stress during access to the memory cells, since the not addressed sectors are disconnected completely from the addressed sector.

[0030] The metal line forming the global word line 35 no longer needs to be arranged between the rows, since each global word line 35 is connected to more than one row (to two rows, in the example shown), and each row is addressed by a polysilicon line (line 57 in figure 5). Thereby, the reliability of the device and the production performance are increased; the line time constant is acceptable; since the length of the local (polysilicon) word lines 36 is limited to that of the sector.

[0031] Each sector 31 is completely disconnected from the others, both horizontally and vertically. Thereby it is possible to completely decouple the faulty sectors or individual lines (rows or columns), and to replace them completely by redundancy sectors or rows, thus overcoming the limitations imposed by the present solutions for redundancy of rows and/or columns. Indeed, in the present architectures, for example in hierarchical column decoding, the presence of a faulty row causes redundancy of all the equivalent rows in the adjacent sectors.

[0032] The possibility of arranging the reading stages 47 in the central channel, rather than at the end of the array, reduces the capacity of the bit lines, and is the more advantageous, the more reading stages 47 are used; in particular, in case of burst devices (in which several words are read in parallel), an architecture of this type is highly advantageous. This solution can advantageously be applied in particular with large size memories, which have many small sectors, since specifically in that case complete division of the memory array is necessary.

[0033] Finally, it is apparent that many modifications and variations can be made to the described and illustrated memory device, all of which come within the scope of the invention, as defined in the attached claims. In particular, the same solution of double hierarchical decoding can be applied not only to the case of row and column decoding, but also to the case of biasing the well accommodating each sector (ip-well or insulated p-well), and all the parts to be biased in an independent manner, when any operation is carried out on a sector.

[0034] In addition, although figure 4 shows a plurality of global row decoders 32, one for each sector row, this arrangement is equivalent to the use of a single global row decoder connected to all the global word lines; similarly, each local row decoder 33 associated with each sector 31 can be provided on a single side of the respective s ctor, if allowed by the spaces and the layout of

the device.

Claims

1. A nonvolatile memory device (30), comprising a memory array (30) including a plurality of cells (50) grouped into sectors (31) and arranged on sector rows and columns, characterised by a double hierarchical decoding.

2. A device according to claim 1, characterised by both hierarchical row decoding, and hierarchical column decoding.

3. A device according to claim 2, characterised in that said memory array (30) has an own width and an own height, and each sector (31) has an own width and an own height;
and in that it comprises:

   at least one global row decoder (32) and one global column decoder (41);
   a plurality of global word lines (35), extending substantially along the entire length of said memory array (30) from said global row decoder;
   a plurality of global bit lines (42) extending along a plurality of sectors (31), in a height direction of said memory array (30) from said global column decoder (41);
   a plurality of local row decoders (33), one for each sector (31);
   a plurality of local column decoders (40), one for each sector;
   a plurality of local word lines (36) extending substantially along the entire width of each said sector (31) from a respective local row decoder (33);
   a plurality of local bit lines (43) extending substantially along the entire height of each sector from a respective local bit decoder (40);
   each global word line (35) being connected to a plurality of local word lines (36) aligned with each other; and each global bit line (42) being connected to a plurality of local bit lines (43) aligned with each other.

4. A device according to claim 3, characterised in that each global word line (35) is connected to at least two local word lines (36) for each sector (31), the local word lines (36) of adjacent sectors (31) connected to a same global word line (35) being aligned along at least two adjacent rows of cells (50).

5. A device according to claim 3 or claim 4, characterised in that each global bit line (42) is connected to at least two local bit lines (43) for each sector (31), the local bit lines (43) of adjacent sectors (31) connected to a same global bit line (42) being aligned along at least two adjacent columns of cells (50).

6. A device according to any of claims 3-5, characterised in that said global column decoder (41) is arranged centrally to said memory array (30) and separates from each other an upper half (30a) and a lower half (30b) of said memory array (30).

7. A device according to claim 6, characterised in that a first plurality of global bit lines (42) extends in said upper half (30a), and a second plurality of global bit lines (42) extends in said lower half (30b) of said memory array (30); each global bit line (42) of said first plurality of global bit lines being aligned with a respective global bit line of said second plurality of global bit lines; and said local bit lines (43) connected to said global bit lines of said first and second plurality being aligned with each other along columns of cells (50).

8. A device according to claim 6 or claim 7, characterised by a reading stage (47) arranged centrally inside said memory array (30), between said upper half (30a) and said lower half (30b) of said memory array (30).

9. A device according to any of claims 3-8, characterised by metal layers; a lower metal layer (61) defining said local bit lines (43); an intermediate metal layer (64) defining said global rows (35), and an upper metal layer (67) defining said global bit lines (42).

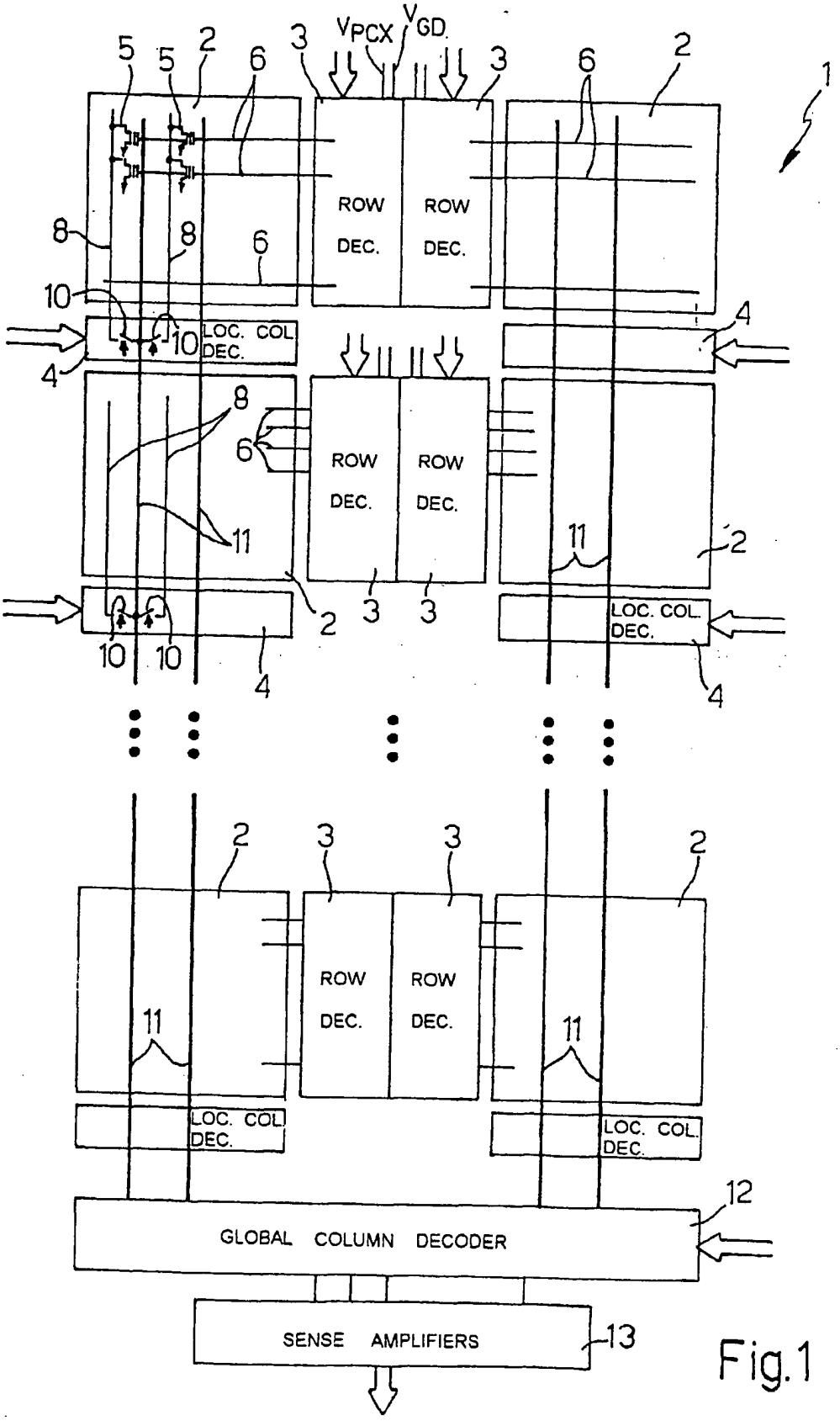
VPCX
VGD
ROW DEC.
ROW DEC.
LOC. COL DEC.
GLOBAL COLUMN DECODER
SENSE AMPLIFIERS

Fig.1

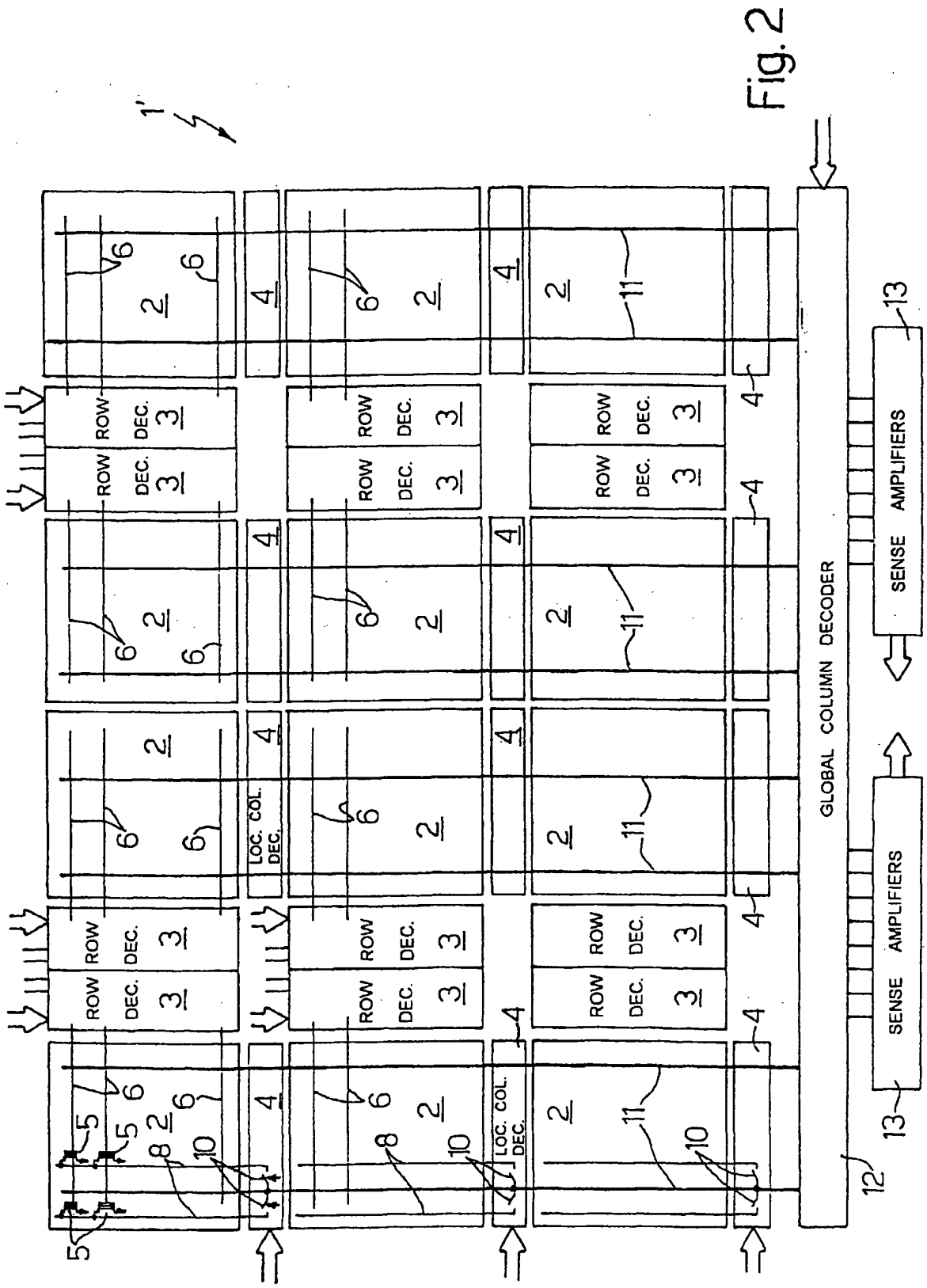
1'
ROW DEC.
LOC. COL. DEC.
GLOBAL COLUMN DECODER
SENSE AMPLIFIERS
SENSE AMPLIFIERS

Fig. 2

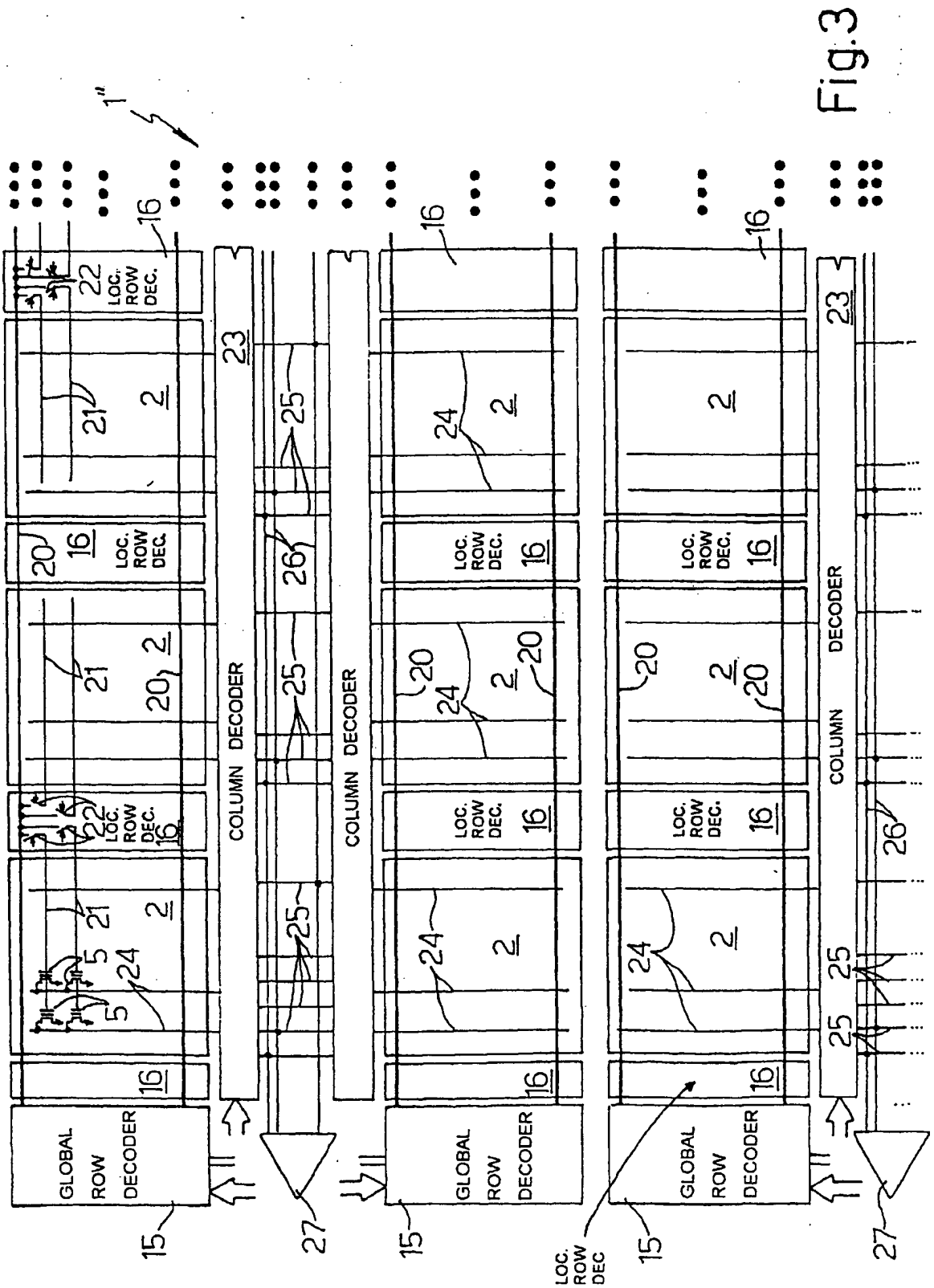
1''
GLOBAL ROW DECODER
COLUMN DECODER
LOC. ROW DEC.
LOC. ROW DEC
COLUMN DECODER
15
16
20
21
22
23
24
25
26
27
2
5

Fig.3

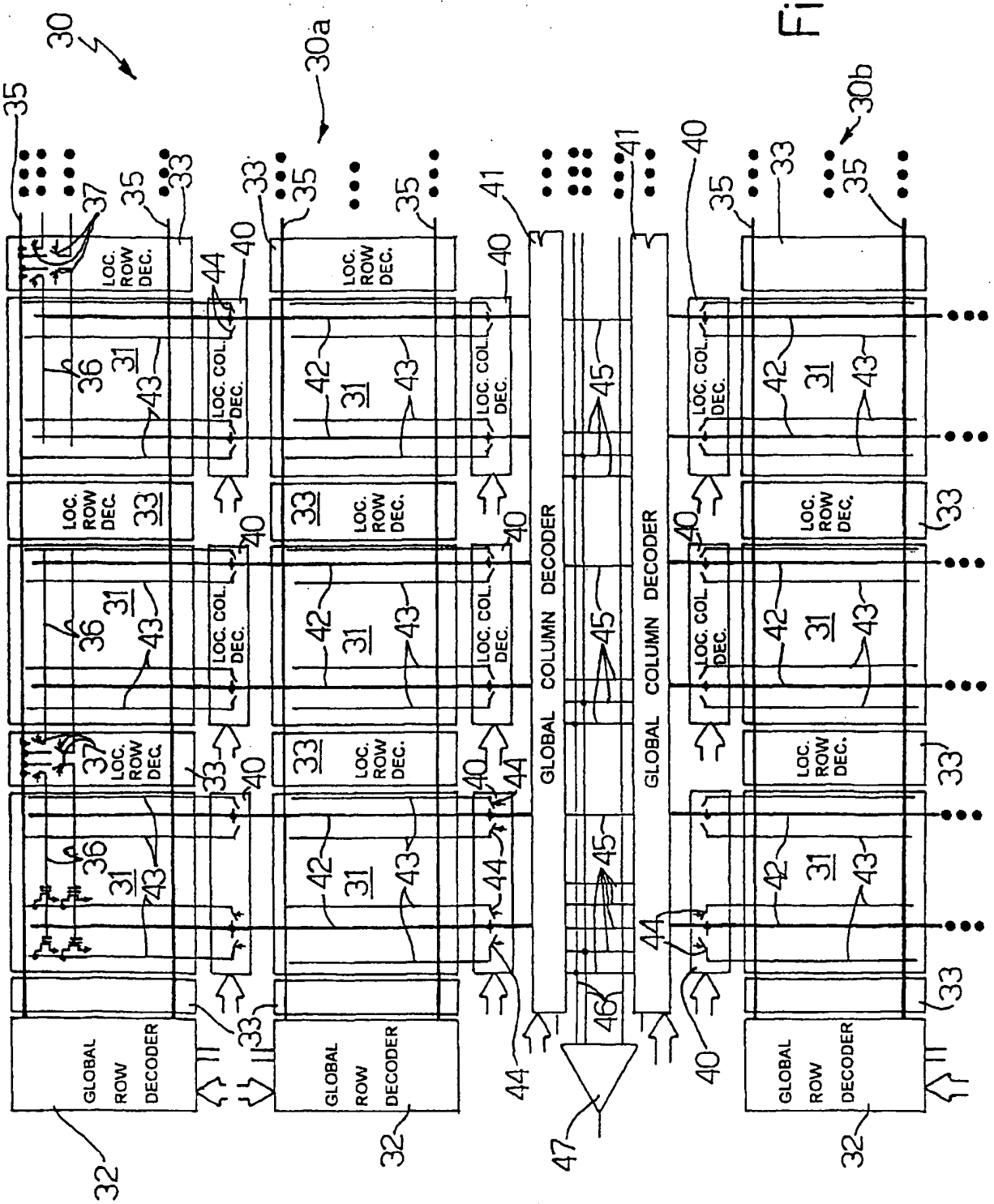
30
30a
30b
GLOBAL ROW DECODER
LOC. ROW DEC.
LOC. COL. DEC.
GLOBAL COLUMN DECODER
31
32
33
35
36
37
40
41
42
43
44
45
46
47

Fig. 4

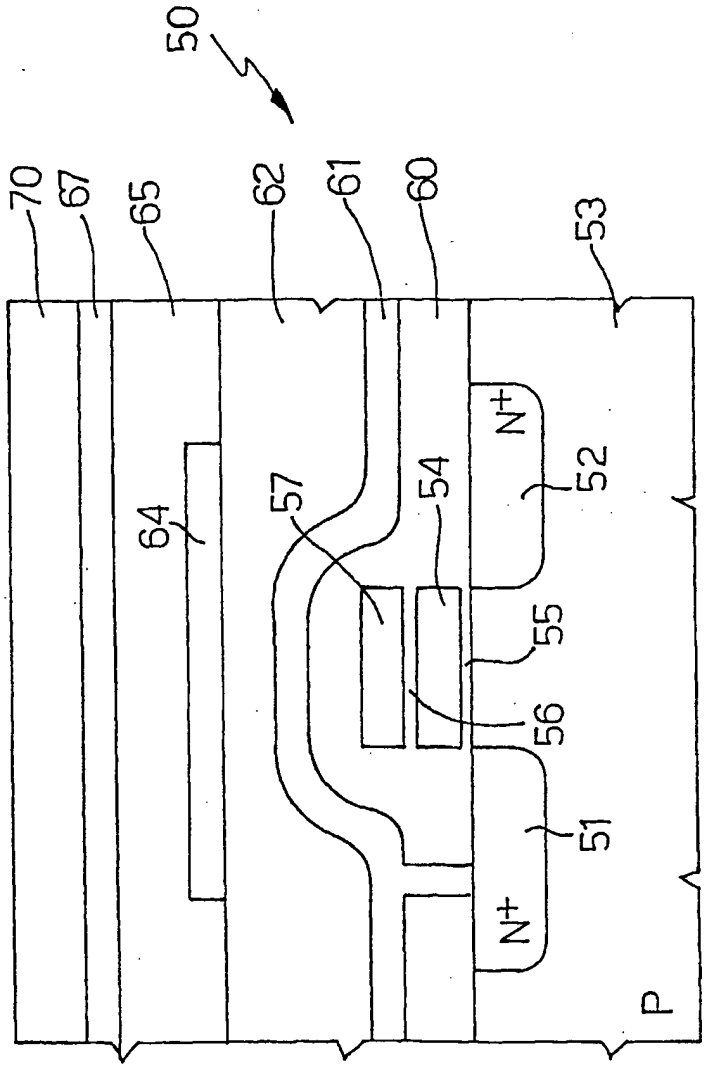

Fig. 5

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 99 83 0236

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION
X	US 5 854 770 A (PASCUCCI LUIGI) 29 December 1998 (1998-12-29)	1,2	G11C8/00 G11C16/06
A	* column 2, line 19 - column 4, line 16 * * column 4, line 52 - line 65 * * figures 4A-4D *	3,6	G11C7/18
X	US 5 663 903 A (GUO JENG-JONG) 2 September 1997 (1997-09-02)	1,2	
A	* column 1, line 54 - column 2, line 9 * * figure 3 *	3	
A	EP 0 834 881 A (SGS THOMSON MICROELECTRONICS) 8 April 1998 (1998-04-08) * column 1, line 1 - column 3, line 6 * * figures 3,4 *	1,2	
A	CHEN T ET AL: "OPTIMIZATION OF THE NUMBER OF LEVELS OF HIERARCHY IN LARGE-SCALE HIERARCHICAL MEMORY SYSTEMS" PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, SAN DIEGO, MAY 10 - 13, 1992, vol. 5, no. CONF. 25, 10 May 1992 (1992-05-10), pages 2104-2107, XP000338403 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERSISBN: 0-7803-0593-0 * page 2105, column 1, paragraph 2 - page 2106, column 1, paragraph 2 * * figure 2 *	1,2	

TECHNICAL FIELDS SEARCHED
G11C

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	3 September 1999	Colling, P

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 99 83 0236

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

03-09-1999

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5854770 A	29-12-1998	NONE	
US 5663903 A	02-09-1997	NONE	
EP 0834881 A	08-04-1998	NONE	